DOCUMENT RESUME

ED 208 024 TM 810 715

AUTHOR Hedges, Larry V.
TITLE Statistical Aspects of Effect Size Estimation.
SPONS AGENCY Spencer Foundation, Chicago, Ill.
PUB DATE Apr 81
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Literature Reviews; *Mathematical Models; Maximum Likelihood Statistics; *Statistical Analysis
IDENTIFIERS *Effect Size; Estimation (Mathematics); *Meta Analysis; Sample Size

ABSTRACT

When the results of a series of independent studies are combined, it is useful to quantitatively estimate the magnitude of the effects. Several methods for estimating effect size are compared in this paper. Glass' estimator and the uniformly minimum variance unbiased estimator are based on the ratio of the sample mean difference and the pooled within-group standard deviation. The third estimator is the maximum likelihood estimator. The fourth estimator is a shrunken form of the minimum variance unbiased estimator. All four estimators are shown to be equivalent in large samples, but they differ in finite samples. Two procedures for testing the fit of the data to the proposed structural model and for detection of outliers, an example of the application of the techniques, and a summary of recommendations on statistical procedures for estimation of effect size from a series of experiments are presented. (Author/BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED208024

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

STATISTICAL ASPECTS OF EFFECT SIZE ESTIMATION[1]

by

Larry V. Hedges

The University of Chicago

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
L. V. Hedges
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the 1981 Annual Meeting of the American Educational Research Association

TM 810 715

1. This research was supported by the Spencer Foundation

ESTIMATION OF EFFECT SIZE FROM A SERIES OF EXPERIMENTS

1. Introduction

A number of authors have recently shown an interest in empirical methods of combining the results of a series of independent studies. Glass (1976) was among the first authors to call for the use of quantitative procedures in research integration as a supplement to the discursive review. Examples of the use of these techniques are the review of psychotherapy outcome studies (Smith & Glass, 1977) and the review of class size literature (Glass & Smith, 1979). Both examples demonstrate the utility of using quantitative estimates of the magnitude of effects in research reviews.

Several methods for estimating effect size are compared in this paper. A precise definition of effect size, a structural model for a series of experiments, and a conceptual rationale for the model are presented in Section 2. Methods for estimation of effect size from a single experiment are discussed in Section 3. Section 4 is a presentation of results on estimation of effect size from a series of experiments when the sample sizes of the individual experiments are all large. Some results on estimation of effect size from a large number of small experiments are presented in Section 5. Two procedures for testing the fit of the data to the proposed structural model and for detection of outliers are presented in Section 6. Section 7 is an example of the application of the techniques presented in this paper. Section 8 is a summary of recommendations on statistical procedures for estimation of effect size from a series of experiments.

2. A Structural Model for a Series of Experiments

The statistical properties of procedures for combining results from a series of experiments depend on the structural model for the results of the experiments. The structural model used in this paper requires that each experiment uses a response scale from a collection of congeneric measures, i.e., each response scale is a linear transformation of a response scale with unit variance within groups. For example, this assumption is satisfied if the tests are linearly equatable.

Conceptually, we assume that each experiment to be combined is a replication of the others, differing only in the response scale and sample sizes. That is, the population value of the treatment effect from each study would be identical if transformed to a common response scale. More precisely, let Y^E_{ij} and Y^C_{ij} be the j^{th} scores on the i^{th} experiment from the experimental and control groups, respectively. Assume that for fixed i, Y^E_{ij} and Y^C_{ij} are normally distributed with means μ^E_i and μ^C_i and common variance β^2_i, i.e.,

$$Y^E_{ij} \sim N(\mu^E_i, \beta^2_i) \quad , \quad j = 1,\ldots,n^E_i \quad , \quad i = 1,\ldots,k \quad ,$$

and

$$Y^C_{ij} \sim N(\mu^C_i, \beta^2_i) \quad , \quad j = 1,\ldots,n^C_i \quad , \quad i = 1,\ldots,k \quad ,$$

where $(\mu^E_i - \mu^C_i)/\beta_i = \delta_i, 1,\ldots,k$. Assume also that $\beta_i > 0$ and that $\delta_i = \delta$, $i = 1,\ldots,k$. Thus δ, the standardized mean difference, is the treatment effect (mean difference) when the response scale has unit variance. The parameter δ is called the effect size.

A more compact representation of the model presented above involves the explicit use of δ, the within group standard deviation β_i, a location (scale mean) parameter γ_i, and a residual term ε. This structural model can be written as

$$Y_{ij}^{E} = \beta_i\delta + \beta_i\gamma_i + \varepsilon_{ij}^{E} \quad , \quad j = 1,\ldots,n_i^{E} \quad , \quad i = 1,\ldots,k \quad ,$$

(1)

$$Y_{ij}^{C} = \beta_i\gamma_i + \varepsilon_{ij}^{C} \quad , \quad j = 1,\ldots,n_i^{C} \quad , \quad i = 1,\ldots,k \quad ,$$

where $\varepsilon_{ij}^{E} \sim N(0,\beta_i^2)$ and $\varepsilon_{ij}^{C} \sim N(0,\beta_i^2)$.

3. Estimation of Effect Size from a Single Experiment

In this section we consider four estimators of the effect size δ based on a single experiment. Two estimators (Glass' estimator and the uniformly minimum variance unbiased estimator) are based on the ratio of the sample mean difference and the pooled within group standard deviation. The third estimator is the maximum likelihood estimator. The fourth estimator is a shrunken form of the minimum variance unbiased estimator. All four estimators are shown to be equivalent in large samples, although they differ in finite samples.

3.1 Glass's Estimator of Effect Size Based on the Standardized Mean Difference

Glass (1976) proposed an estimator of δ based on the sample value of the standardized mean difference for each experiment, which is then averaged in a set of experiments to obtain the estimator based on the series of k experiments.

Let $\bar{Y}_i^E$ and $\bar{Y}_i^C$ be the experimental and control group means for the i^{th} experiment and S_i^C is the sample standard deviation of the control group of the i^{th} experiment, and define:

$$g_i' = \frac{\bar{Y}_i^E - \bar{Y}_i^C}{S_i^C} \quad , \quad i = 1,\ldots,k.$$

Glass proposed the use of the standard deviation of the control group to standardize the mean difference. His argument was that pooling two variances could lead to different standardized values of the identical mean difference within an experiment where several treatments were compared to a control. The argument depends on the fact that sample standard deviations of each group will surely differ. In many cases, however, the model of equal population variances is reasonable, which suggests that the most precise estimate of the population variance is obtained by pooling. Model (1) includes only two groups per experiment, with equal population variances, so Glass' argument does not apply. Hence a modified Glass estimator can be obtained by using a pooled estimate of the standard deviation.

If g_i is the modified estimator for the i^{th} experiment, then

$$(2) \qquad g_i = \frac{\bar{Y}_i^E - \bar{Y}_i^C}{S_i} \quad ,$$

where S_i is the pooled standard deviation estimate, i.e.

$$S_i^2 = \frac{(n_i^E - 1)(S_i^E)^2 + (n_i^C - 1)(S_i^C)^2}{n_i^E + n_i^C - 2} ,$$

$\bar{Y}_i^E$, $\bar{Y}_i^C$, $(S_i^E)^2$, $(S_i^C)^2$, n_i^E, and n_i^C are the experimental and control group means, variances and sample sizes respectively in the i^{th} experiment.

Hedges (1981) showed that both g' and g are distributed as $\sqrt{(n_i^E + n_i^C)/n_i^E n_i^C}$ times a noncentral t-variate with noncentrality parameter $\delta\sqrt{n_i^E n_i^C/(n_i^E + n_i^C)}$ and $(n_i^C - 1)$ degrees of freedom for g_i' and $(n_i^E + n_i^C - 2)$ degrees of freedom for g_i. It was further shown that g_i always has smaller bias and variance than g_i'. Therefore g_i' will not be considered further in this paper. For reference we now list two results from Hedges (1981) Theorem 1.

Fact 1: Suppose that g_i is defined as in (2) then the mean and variance of g_i under model (1) are given by

(3a) $$E(g_i) = \delta / c(m_i),$$

(3b) $$\text{Var}(g_i) = \frac{m_i}{(m_i - 2)\tilde{n}_i}[1 + \tilde{n}_i\delta^2] - \delta^2/[c(m_i)]^2,$$

where $m_i = n_i^E + n_i^C - 2$, $\tilde{n}_i = n_i^E n_i^C / (n_i^E + n_i^C)$, and

(4) $$c(m) = \frac{\Gamma(m/2)}{\sqrt{m/2}\ \Gamma[(m-1)/2]} .$$

Fact 2: Suppose that g_i is defined as in (2) then when n_i^E and n_i^C are both large, g_i is approximately normally distributed with mean δ and variance

(5) $$\sigma_i^2(\delta) = \frac{n_i^E + n_i^C}{n_i^E n_i^C} + \frac{\delta^2}{2(n_i^E + n_i^C)} .$$

3.2 Unbiased Estimation of δ

The evaluation of the expectation of g_i reveals that g_i has a small sample bias. Since $E(g) = \delta/c(m)$, it is clear that an unbiased estimator may be obtained. In particular the estimator g_i^U defined by

$$g_i^U = c(m_i)g_i \tag{6}$$

is unbiased, where $c(m)$ is defined by (4). Since $c(m) < 1$, the estimator (g_i^U) also has smaller variance then g_i. Thus g_i^U is generally preferrable to g_i. When $n_i^E = n_i^C$, Hedges (1981) has shown that g_i^U is the unique minimum variance unbiased estimator of δ. The variance of g_i^U is given in the following fact.

Fact 3: Suppose g_i^U is defined as in (6), then g_i^U is an unbiased estimator of δ under model (1) with variance given by

$$\text{Var}\ (g_i^U) = \frac{m_i\,[c(m_i)]^2}{(m_i - 2)\tilde{n}_i}\ [1 + \tilde{n}_i\delta^2\ ] - \delta^2,$$

where $\tilde{n}_i = n_i^E\, n_i^C\ /\ (n_i^E + n_i^C\ )$, and $c(m)$ is given by (4).

Note that g_i^U and g_i both tend to a normal distribution with identical mean and variances in large samples since $c(m) \to 1$ as $m \to \infty$. This implies that g_i^U tends in distribution to g_i in large samples.

3.3 Maximum Likelihood Estimation of Effect Size

Maximum likelihood estimation has the general advantages of consistency and asymptotic efficiency. Maximum likelihood estimation is also widely used even when the requirements of large samples are not met. In Section 4 maximum likelihood estimation is explored for a series of k experiments. In this section we derive the maximum likelihood estimator using the invariance of maximum likelihood estimators.

In the case of a single experiment, the parameter δ is the ratio of the mean difference $(\mu_i^E - \mu_i^C)$ and the within group standard deviation β_i. Thus it follows that the maximum likelihood estimator of δ is the ratio of the maximum likelihood estimators of $(\mu_i^E - \mu_i^C)$ and β_i. These estimators are

$$\text{Est } (\mu_i^E - \mu_i^C) = \bar{Y}^E - \bar{Y}^C$$

and

$$\text{Est } (\beta_i) = \sqrt{\frac{n_i^E + n_i^C}{n_i^E + n_i^C - 2}}\, S,$$

where S^2 is the usual unbiased estimate of the variance. Therefore the maximum likelihood estimator $\hat{\delta}_i$ of δ based on a single experiment is given by

$$(6) \qquad \hat{\delta}_i = \sqrt{\frac{n_i^E + n_i^C}{n_i^E + n_i^C - 2}}\ \frac{\bar{Y}^E - \bar{Y}^C}{S} = g\sqrt{\frac{n_i^E + n_i^C}{n_i^E + n_i^C - 2}}\ .$$

Note that the maximum likelihood estimator is biased in small samples. It is also clear that $\hat{\delta}_i$ has larger absolute bias than g, the Glass estimator. Since $\sqrt{(n_i^E + n_i^C) / (n_i^E + n_i^C - 2)} > 1$, the maximum likelihood estimator

$\hat{\delta}$ has larger variance than g^U. Note that the constant $\sqrt{(n_i^E + n_i^C) / (n_i^E + n_i^C - 2)}$ tends to one in large samples, which implies that the maximum likelihood estimator tends in distribution to the estimators g_i and g_i^U. Thus all three estimators of effect size from a single experiment are equivalent in large samples.

3.4 Shrunken Estimators of Effect Size

It is well known that the minimum variance unbiased estimator may not be the minimum mean squared error estimator. For example, Stein (1956) showed that the mean squared error of the maximum likelihood estimator of the mean of the multivariate normal distribution can be reduced by "shrinking" the estimates toward zero. The principle in these shrunken estimates is that the increase in the bias term of the mean squared error is more than compensated for in the reduction of the variance term of the mean squared error. In this section we show that the minimum variance unbiased estimator is inadmissable under quadratic loss by exhibiting a shrunken estimator with uniformly smaller mean squared error.

Theorem 1: Suppose g_i^U is defined as in (6), then the estimator g_i^S has uniformly smaller mean squared error than g_i^U under model (1) where

$$g_i^S = a(m_i)\, g_i^U , \tag{7}$$

and

$$a(m) = \frac{m-2}{m\,[c(m)]^2} , \tag{8}$$

where $c(m)$ is given by (4).

Proof: Consider an estimator of the form ag^U, where a is a constant. If δ is the true effect size, the mean squared error of ag^U is given by

$$MSE\,(ag^U) = a^2 Var(g^U) + \delta^2\,(1 - a)^2 .$$

The variance of g^U is of the form $c_1 + c_2\,\delta^2$, where c_1 and c_2 are constants, therefore

$$\text{MSE}\,(ag^U) = a^2 c_1 + [\,a^2\,(c_2 + 1) - 2a + 1\,]\,\delta^2.$$

Define the function f(a)

$$f(a) = \text{MSE}\,(ag^U) - \text{MSE}\,(g^U),$$

i.e. f(a) is the difference in mean squared errors of the estimators ag^U and g^U. An explicit expression for f(a) is

$$f(a) = c_1\,(a^2 - 1) + [\,a^2\,(c_2+1) - 2a + 1 - c_2\,]\,\delta^2.$$

Examination of the derivatives of f shows that f has a minimum at

$$a_o = \frac{\delta^2}{c_1 + (c_2 + 1)\,\delta^2}\ .$$

This minimum a_o depends on δ so we note that as $\delta \to \infty$, $a_o \to 1/(c_2 + 1)$. Inserting $a = 1/(c_2 + 1)$ into the expression for f(a) yields

$$\frac{-c_1 c_2 (c_2 + 2) - (c_2^2 + c_2^3)\,\delta^2}{c_2^2 + 2c_2 + 1}\ .$$

In particular, $c_1 > 0$ and $c_2 > 0$ imply that the value of f given above is negative and therefore $\text{MSE}\,(ag^U) < \text{MSE}\,(g^U)$. The value of c_2 given in Fact 3

$$c_2 = \frac{m\,[\,c(m)\,]^2}{(m-2)} - 1,$$

is then inserted into the formula $a = 1/(c_2 + 1)$, which gives the constant a(m).

Table 1

Variance, Bias, and Mean Squared Error of Four Estimators of Effect Size for Various Sample Sizes (m + 2) and Effect Sizes δ

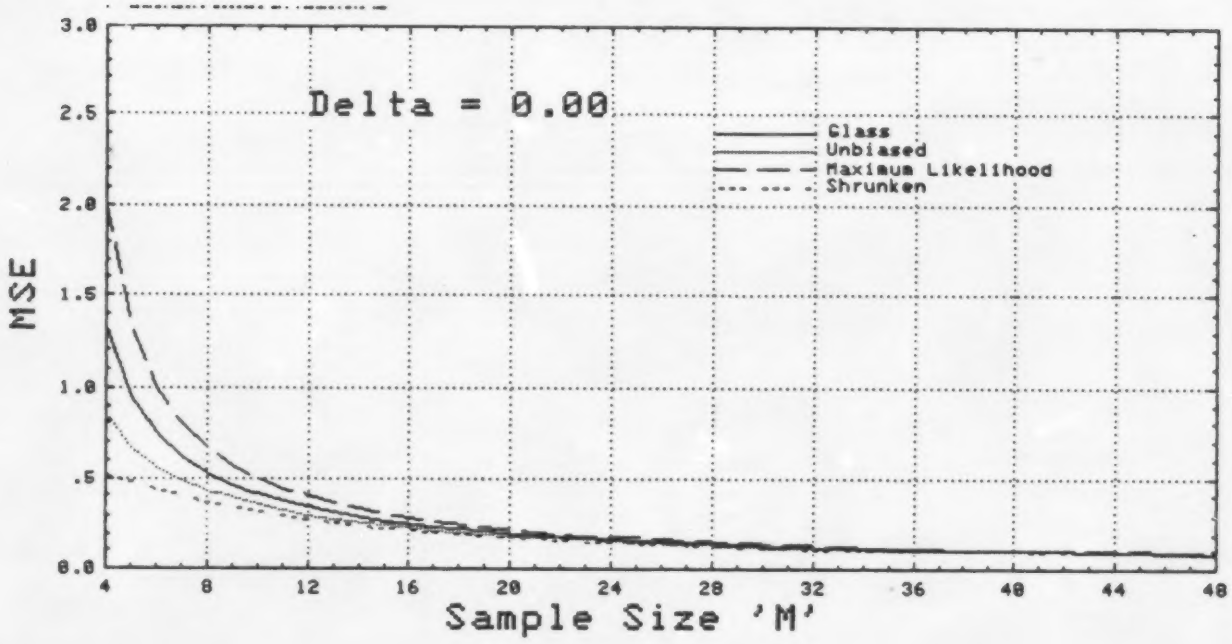
	VARIANCE				BIAS				MSE			
m	$\hat{\delta}$	g	g^U	g^S	$\hat{\delta}$	g	g^U	g^S	$\hat{\delta}$	g	g^U	g^S
					$\delta = 0$							
4	2.000	1.333	.849	.524	0.000	.000	.000	.000	2.000	1.333	.849	.524
6	1.000	.750	.566	.442	.000	.000	.000	.000	1.000	.750	.566	.442
8	.667	.533	.435	.368	.000	.000	.000	.000	0.667	.533	.435	.368
10	.500	.417	.355	.313	.000	.000	.000	.000	.500	.417	.355	.313
15	.308	.271	.244	.226	.000	.000	.000	.000	.308	.271	.244	.226
20	.222	.202	.187	.177	.000	.000	.000	.000	.222	.202	.187	.177
25	.174	.161	.151	.145	.000	.000	.000	.000	.174	.161	.151	.145
50	.083	.080	.078	.076	.000	.000	.000	.000	.083	.080	.078	.076
100	.041	.040	.039	.039	.000	.000	.000	.000	.041	.040	.039	.039
					$\delta = .2$							
4	2.026	1.351	.860	.530	0.107	.051	.000	-.043	2.037	1.353	0.860	0.532
6	1.009	.757	.571	.446	.066	.030	.000	-.023	1.014	.758	.571	.446
8	.672	.538	.438	.371	.048	.022	.000	-.016	.674	.538	.438	.371
10	.504	.420	.357	.315	.037	.017	.000	-.012	.505	.420	.357	.316
15	.310	.273	.246	.228	.024	.011	.000	-.008	.310	.273	.246	.228
20	.224	.203	.188	.178	.018	.008	.000	-.005	.224	.203	.188	.178
25	.175	.162	.152	.146	.014	.006	.000	-.004	.175	.162	.152	.146
50	.084	.081	.078	.076	.007	.003	.000	-.002	.084	.081	.078	.076
100	.041	.040	.039	.039	.004	.002	.000	-.001	.041	.040	.039	.039

Table 1 (cont.)

m	VARIANCE $\hat{\delta}$	VARIANCE g	VARIANCE g^U	VARIANCE g^S	BIAS $\hat{\delta}$	BIAS g	BIAS g^U	BIAS g^S	MSE $\hat{\delta}$	MSE g	MSE g^U	MSE g^S
						$\delta = .5$						
4	2.161	1.441	.917	.566	.267	.127	.000	-.107	2.233	1.457	.917	.577
6	1.058	.794	.599	.468	.165	.076	.000	-.058	1.085	.799	.599	.471
8	.670	.560	.456	.386	.119	.054	.000	-.040	.714	.563	.456	.388
10	.523	.436	.371	.327	.094	.042	.000	-.030	.531	.437	.371	.328
15	.320	.282	.254	.235	.061	.027	.000	-.019	.324	.283	.254	.236
20	.231	.210	.194	.184	.045	.020	.000	-.014	.233	.210	.194	.184
25	.180	.167	.157	.150	.036	.016	.000	-.011	.182	.167	.157	.150
50	.087	.083	.080	.079	.018	.008	.000	-.005	.086	.083	.083	.079
100	.042	.041	.041	.040	.009	.004	.000	-.002	.042	.041	.041	.040
						$\delta = 1.0$						
4	2.644	1.763	1.122	0.692	.535	.253	.000	-.215	2.930	1.827	1.122	.738
6	1.233	.925	.698	.545	.329	.151	.000	-.116	1.341	1.240	.698	.558
8	.799	.639	.521	.441	.239	.108	.000	-.080	.856	.770	.521	.448
10	.591	.492	.419	.370	.187	.084	.000	-.060	.626	.565	.419	.374
15	.357	.315	.284	.263	.122	.054	.000	-.038	.372	.448	.284	.264
20	.256	.232	.215	.203	.090	.040	.000	-.027	.264	.318	.215	.204
25	.199	.184	.173	.166	.072	.031	.000	-.022	.204	.185	.173	.166
50	.094	.090	.088	.086	.054	.015	.000	-.010	.096	.091	.088	.086
100	.046	.045	.044	.044	.018	.008	.000	-.005	.046	.045	.044	.044

Figure 1

Exact Mean Squared Error of Four Estimators of Effect Size as a Function of Sample Size (Total Degrees of Freedom) M.



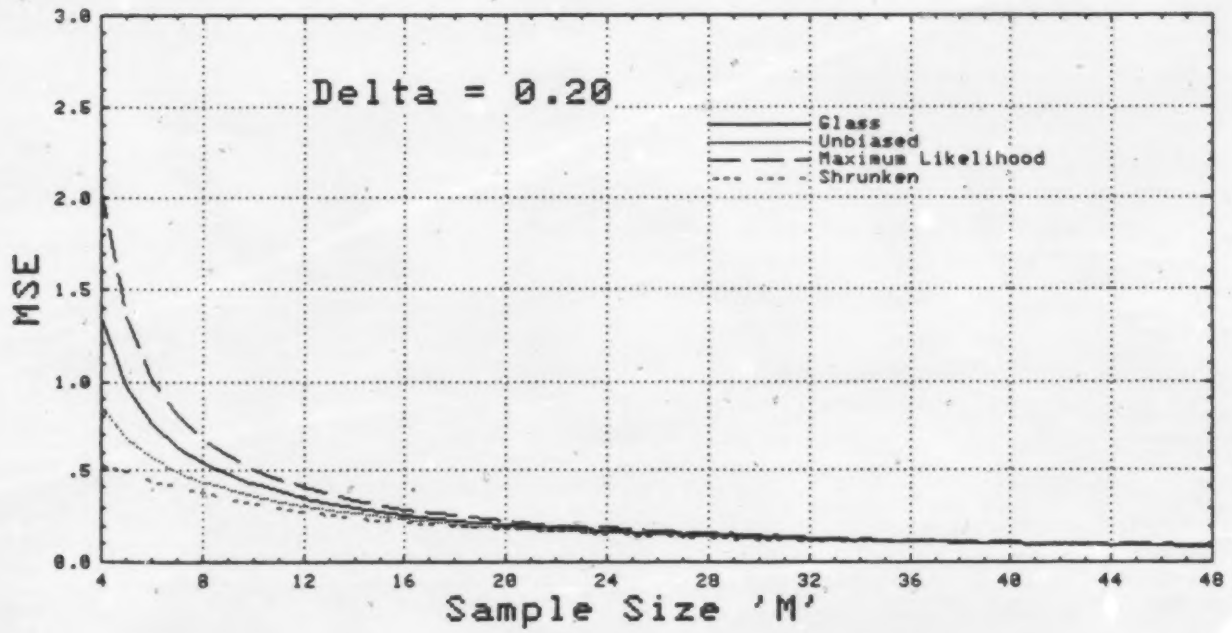
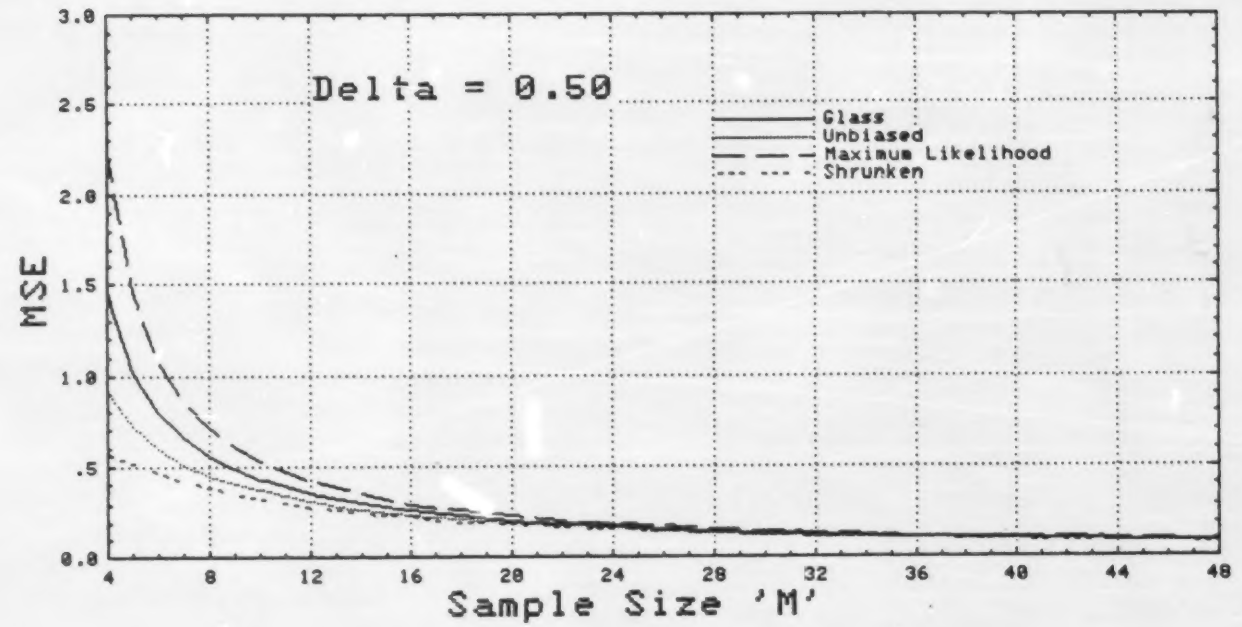
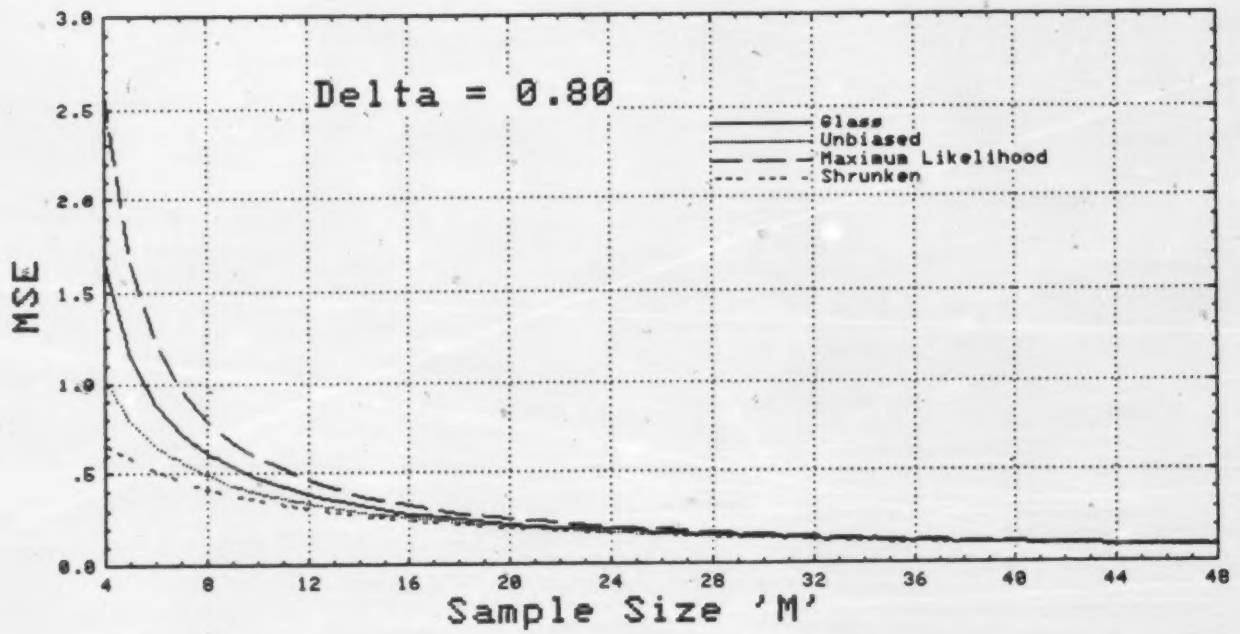


Figure 2

Exact Mean Squared Error for Four Estimators of Effect Size as a Function of Sample Size (Total Degrees of Freedom) M.





3.5 Comparing Estimators of Effect Size

Four estimators of the effect size δ were examined in Sections 3.1 to 3.4. Each of these estimators has the form of a constant times g, where the constant tends to one as the sample size tends to infinity. Thus the four estimators are equivalent in large samples, i.e. they have the same asymptotic distribution. The estimators differ in finite samples, however. The estimators produce estimates that are ordered in absolute value, i.e.,

$$|\hat{\delta}| \geq |g| \geq |g^U| \geq |g^S|,$$

where equality holds only if all the estimates are zero. The same ordering holds with respect to the variances of the estimators,

$$\mathrm{Var}(\hat{\delta}) > \mathrm{Var}(g) > \mathrm{Var}(g^U) > \mathrm{Var}(g^S).$$

Note that $\hat{\delta}$ and g have positive bias in that $|E(\hat{\delta})| - |\delta|$ and $|E(g)| - |\delta|$ are positive, when $\delta \neq 0$ while g^S has negative bias since $|E(g^S)| - |\delta|$ is negative when $\delta \neq 0$. Table 1 is a tabulation of the variance, bias, and mean squared error of the four estimators for selected sample sizes and effect sizes δ. It is clear that g^S is the best estimator by the criterion of mean squared error, and that the differences are largest when the sample size is small. Figures 1 and 2 present graphical illustrations of the mean squared error of the four estimators as a function of sample size.

4. Estimation of Effect Size from a Series of Experiments with Large Samples

In this section we consider estimation of effect size from a series of experiments where the sample size of each experiment is large. Two estimators are considered, one estimator is based on linear combinations of effect size estimates g_i from each experiment. The other estimator is the maximum likelihood estimator based on a series of experiments.

4.1 An Estimator Based on a Linear Combination of g_i's

The sample size of each experiment is assumed to be large in this section. Therefore the estimator g_i from each experiment will not differ too much from g_i^U. For this reason we will not consider weighted estimates of g_i and g_i^U separately. The weighted estimator G based on a series of experiments is given by

$$G = \sum_{i=1}^{k} w_i g_i \,, \tag{9}$$

where $w_i > 0$, $i = 1, \ldots k$ and $\sum_{i=1}^{k} w_i = 1$.

The weights w_i which minimize the variance of G are given by

$$w_i = \frac{1/v_i}{\sum_{j=1}^{k} 1/v_j} \,, \quad i = 1,\ldots k, \tag{10}$$

where $v_i = \mathrm{Var}\,(g_i)$, $i = 1,\ldots k$.

Since the weights w_i depend on the parameter δ via the variances v_i there is in general no a priori expression for the weights. One method of estimating the optimal weights is to use g_i as an estimate of δ for calculating each v_i. These estimated weights are given below.

Theorem 2: Suppose that the sample sizes n_i^E and n_i^C in each experiment are large, then if G is the weighted estimator defined in (9), then estimates of the weights w_i which minimize the variance of G are

$$\hat{w}_i = \frac{1/\hat{v}_i}{\sum_{j=1}^{k} (1/\hat{v}_j)} \quad , i = 1,,,,k,$$

where

$$\hat{v}_i = \frac{n_i^E + n_i^C}{n_i^E n_i^C} + \frac{g_i^2}{2(n_i^E + n_i^C)} \quad , i = 1,\ldots k.$$

Formally, if $N = \sum_{i=1}^{k} (n_i^E + n_i^C)$, $\pi_i^E = n_i^E/N$, $\pi_i^C = n_i^C/N$, $i = 1,\ldots k$,

then if π_i^E and π_i^C, $i = 1,\ldots k$ remain fixed as $N \to \infty$, $\sqrt{N}\,(G - \delta)$ has an asymptotic normal distribution with mean zero and variance

$$\sigma^2 = \frac{1}{\sum_{i=1}^{k} \frac{2\,\pi_i^E \pi_i^C (\pi_i^E + \pi_i^C)}{2\,(\pi_i^E + \pi_i^C)^2 + \pi_i^E \pi_i^C \delta^2}} .$$

Informally, G has an asymptotic normal distribution with mean δ and variance

$$\text{Var } (G) = \frac{1}{\sum_{i=1}^{k} \frac{2n_i^E n_i^C (n_i^E + n_i^C)}{2(n_i^E + n_i^C)^2 + n_i^E n_i^C \delta^2}}$$

Proof: The proof is a straightforward application of the asymptotic theory for weighted estimators (see e.g. Rao, 1965, p.398-9).

Corollary 1: When sample sizes n_i^E and n_i^C, $i = 1,\ldots k$, are large, $n_i^E = n_i^C = n_i'$, $i = 1,\ldots k$, and G is the weighted estimator (4), the weights w_i which minimize the variance of G are given by

$$w_i = \frac{n_i'}{\sum_{j=1}^{k} (n_j')} ,$$

and G has an asymptotic normal distribution with mean δ and variance

$$\text{Var } (G) = \frac{2 + \delta^2/4}{\sum_{i=1}^{k} (n_i')} .$$

4.2 Maximum Likelihood Estimation of δ Based on a Series of Experiments

The log likelihood L for the i^{th} experiment under model (1) is

$$(11) \qquad L_i = 1/2 \log (2\pi)^{n_i} - n_i \log \beta_i + \\ - \frac{1}{2\beta_i^2} \left[\sum_{j=1}^{n_i^E} (Y_{ij}^E - \beta_i\delta - \beta_i\gamma_i)^2 + \sum_{j=1}^{n_i^C} (Y_{ij}^C - \beta_i\gamma_i)^2 \right] ,$$

where $n_i = n_i^E + n_i^C$, $i = 1,\ldots k$.

Setting the derivatives of $L = \sum_{i=1}^{k} L_i$ equal to zero yields a system of $2k + 1$ likelihood equations:

$$(12a) \qquad \frac{\partial L}{\partial \gamma_i} = \frac{(n_i^E \bar{Y}_i^E + n_i^C \bar{Y}_i^C)}{\beta_i} - n_i^E \delta - (n_i^E + n_i^C)\gamma_i = 0, \; i = 1,\ldots k \; ,$$

$$(12b) \qquad \frac{\partial L}{\partial \beta_i} = -\frac{(n_i^E + n_i^C)}{\beta_i} +$$

$$\frac{1}{\beta_i^3}\left[\sum_{j=1}^{n_i^E} (Y_{ij}^E)^2 + \sum_{j=1}^{n_i^C} (Y_{ij}^C)^2 - n_i^E \bar{Y}_i^E \delta\beta_i - (n_i^E \bar{Y}_i^E + n_i^C \bar{Y}_i^C)\beta_i\gamma_i\right]$$

$$= 0 \; , \quad i = 1,\ldots k \; ,$$

$$(12c) \qquad \frac{\partial L}{\partial \delta} = \sum_{i=1}^{k} n_i^E \left[\frac{\bar{Y}_i^E}{\beta_i} - \delta - \gamma_i\right] = 0 \; .$$

The $2k$ likelihood equations (12a) and (12b) can be solved simultaneously to obtain expressions for β_i and γ_i, $i = 1,\ldots k$, in terms of δ and sample data.

Substitution of these expressions into (12c) yields a single equation involving δ and sample data from which estimates can be obtained numerically. (A computer program is available from the author). The resulting equation is of the form

$$(13) \qquad A\delta + \sum_{i=1}^{k} B_i \sqrt{\delta^2 + C_i} = 0 \; ,$$

where $A = \sum_{i=1}^{k} n_i^E n_i^C (Z_i - 2) / (n_i^E + n_i^C)$, $B_i = \pm n_i^E n_i^C Z_i / (n_i^E + n_i^C)$ and the sign is determined by the sign of $(\bar{Y}_i^E - \bar{Y}_i^C)$, $C_i = 4(n_i^E + n_i^C)^2 / n_i^E n_i^C Z_i$, and Z_i is the ratio of the between-group sum of squares and the total sum of squares about the mean of the i^{th} experiment.

In general, when k >2, equation (13) is equivalent to a polynomial of degree greater than three, so closed form solutions cannot be obtained. Since the equation is well behaved, algorithms for the numerical solution of (13) are easy to implement on a computer. Once $\hat{\delta}$ is obtained as a solution to (13) the estimates of γ_i and β_i are given by

$$(14) \qquad \hat{\beta}_i = \frac{\{-\tilde{n}_i (\bar{Y}_i^E - \bar{Y}_i^C)\hat{\delta} + \sqrt{\tilde{n}_i \hat{\delta}^2 (\bar{Y}_i^E - \bar{Y}_i^C)^2 + 4n_i SSY_i}\}}{2n_i}, \quad i = 1,\ldots k,$$

where SSY_i is the total sum of squares about the mean for the i^{th} experiment, and

$$(15) \qquad \hat{\gamma}_i = \frac{n_i^E \bar{Y}_i^E + n_i^C \bar{Y}_i^C}{\hat{\beta}_i (n_i^E + n_i^C)} - \frac{n_i^E \hat{\delta}}{n_i^E + n_i^C}, \quad i = 1,\ldots k.$$

Although the likelihood equations (12a, b, c) do not lead to closed form expressions for the maximum likelihood estimate of δ, the asymptotic covariance structure can still be derived. The asymptotic variance of the maximum likelihood estimator is obtained by inverting the expected value of the matrix of second order partial derivatives of the likelihood function. This procedure yields the asymptotic variance given below.

Theorem 3: Suppose $\hat{\delta}$ is the maximum likelihood estimator of the standardized mean difference δ from model (1), where $N = \sum_{i=1}^{k} (n_i^E + n_i^C)$, and let $\pi_i^E = n_i^E / N$, $\pi_i^C = n_i^C / N$, $i = 1,\ldots k$ remain fixed as $N \to \infty$, then

$$\sqrt{N}(\hat{\delta} - \delta)$$

has an asymptotic normal distribution with mean zero and variance

$$(16) \qquad \sigma^2 = \frac{1}{\sum_{i=1}^{k} \frac{2\pi_i^E \pi_i^C (\pi_i^E + \pi_i^C)}{2(\pi_i^E + \pi_i^C)^2 + \pi_i^E \pi_i^C \delta^2}} .$$

Informally, $\hat{\delta}$ has an asymptotic normal distribution with mean δ and variance

$$(17) \qquad \text{Var}(\hat{\delta}) = \frac{1}{\sum_{i=1}^{k} \frac{2n_i^E n_i^C (n_i^E + n_i^C)}{2(n_i^E + n_i^C)^2 + n_i^E n_i^C \delta^2}}$$

If some of the experimental and control group sample sizes are equal, the expression (17) for the asymptotic variance of the maximum likelihood estimator is considerably simpler. The next corollary states a special case.

Corollary: If $n_i^E = n_i^C = n_i'$, i=1,...k, then informally the asypmtotic variance of the maximum likelihood estimator $\hat{\delta}$ of δ is

$$\text{Var}(\hat{\delta}) = \frac{2 + \frac{\delta^2}{4}}{\sum_{i=1}^{k} n_i'} .$$

Proof of Theorem:

The asymptotic covariance matrix of the maximum likelihood estimates is obtained by inverting minus the expectation of the matrix of second order partial derivatives. The expectations of the second order partial derivatives of the likelihood (11) are:

(18a) $$E\left[\frac{\partial^2 L}{\partial \delta^2}\right] = -N \sum_{i=1}^{k} \pi_i^E ,$$

(18b) $$E\left[\frac{\partial^2 L}{\partial \gamma_i \partial \beta_i}\right] = -N(\pi_i^E + \pi_i^C) , \quad i = 1,\ldots,k$$

(18c) $$E\left[\frac{\partial^2 L}{\partial \beta_i^2}\right] = E\left[\frac{n_i^E + n_i^C}{\beta_i} - \frac{1}{\beta_i^4}\left(\sum_{i=1}^{n_i^E} (Y_i^E)^2 + \sum_{i=1}^{n_i^C} (Y_i^C)^2 + 2n_i^E \bar{Y}_i^E \delta\beta_i + 2(n_i^E \bar{Y}_i^E + n_i^C \bar{Y}_i^C)\beta_i\gamma_i\right)\right]$$

$$= \frac{-N(2[\pi_i^E + \pi_i^C] + \pi_i^E[\delta + \gamma_i]^2 + \pi_i^C\gamma_i^2}{\beta_i^2} , \quad i = 1,\ldots,k ,$$

(18d) $$E\left[\frac{\partial^2 L}{\partial \gamma_i \partial \gamma_j}\right] = E\left[\frac{\partial^2 L}{\partial \beta_i \partial \beta_j}\right] = E\left[\frac{\partial^2 L}{\partial \gamma_i \partial \beta_j}\right] = 0 , \quad i \neq j , \quad i, j = 1,\ldots,k ,$$

(18e) $$E\left[\frac{\partial^2 L}{\partial \gamma_i \partial \beta_i}\right] = E\left[\frac{n_i^E \bar{Y}_i^E + n_i^C \bar{Y}_i^C}{-\beta_i^2}\right] = \frac{-N[\pi_i^E\delta + (\pi_i^E + \pi_i^C)\gamma_i]}{\beta_i} , \quad i = 1,\ldots,k$$

(18f) $$E\left[\frac{\partial^2 L}{\partial \gamma_i \partial \delta}\right] = -N\pi_i^E , \quad i = 1,\ldots,k ,$$

(18g) $$E\left[\frac{\partial^2 L}{\partial \beta_i \partial \delta}\right] = E\left[\frac{-n_i^E \bar{Y}_i^E}{\beta_i^2}\right] = -\frac{N\pi_i^E(\delta + \gamma_i)}{\beta_i} , \quad i = 1,\ldots,k ,$$

where $\bar{Y}_i^E$ and $\bar{Y}_i^C$, $i = 1,\ldots,k$ are the experimental and control group means, respectively.

Inversion of the matrix of expectations is facilitated by ordering the $2k + 1$ parameters according to the vector $\theta = (\beta_1,\ldots,\beta_k, \gamma_1,\ldots,\gamma_k,\delta)$. The matrix Ψ defined by $\psi_{ij} = -E(\partial^2 L/\partial\theta_i\partial\theta_j)$ can then be partitioned into simple submatrices. The submatrix formed by the first k rows and columns of Ψ is a

k x k diagonal matrix. Similarly, the submatrix formed by rows k + 1 to 2k and the columns k + 1 to 2k of Ψ is a k x k diagonal matrix. The other two submatrices composing the remainder of the first 2k rows and columns are also k x k diagonal matrices. Thus the matrix Ψ can be partitioned into

$$\Psi = \begin{bmatrix} D_1 & D_2 & u \\ D_2 & D_3 & v \\ u' & v' & d \end{bmatrix} ,$$

where $D_1 = \text{diag}(a_1,\ldots a_k)$, $D_2 = \text{diag}(b_1,\ldots b_k)$, and $D_3 = \text{diag}(c_1,\ldots c_k)$ are k x k diagonal matrices, $u' = (u_1,\ldots u_k)$ and $v' = (v_1,\ldots v_k)$ are k dimensional row vectors, and d is a scalar.

Therefore Ψ^{-1} is given by

$$\Psi^{-1} = \begin{bmatrix} G^{-1} + \ell h^{-1}\ell' & -\ell h^{-1} \\ -h^{-1}\ell' & h^{-1} \end{bmatrix}$$

where

$$G^{-1} = \begin{bmatrix} D^{-1} + TS^{-1}T' & -TS^{-1} \\ -S^{-1}T' & -S^{-1} \end{bmatrix} ,$$

$$S = D_3 - D_2^{-1} D_1^{-1} D_2, \quad T = D_1^{-1} D_2,$$

$$h = d - [u'v']G^{-1}\begin{bmatrix} u \\ v \end{bmatrix} ,$$

and

$$\ell = G^{-1}\begin{bmatrix} u \\ v \end{bmatrix} .$$

Therefore if $\Psi^{-1} = (\psi^{ij})$,

$$\psi 2k+1,\ 2k+1 = \sum_{i=1}^{k} \frac{d}{k} - \left[\frac{c_i u_i^2 + a_i v_i^2 - 2 b_i u_i v_i}{a_i c_i - b_i^2} \right]^{-1} .$$

Substitution of the expressions from (18) for d, a_i, b_i, c_i, u_i, and v_i, $i = 1,\ldots k$ yields the expression (16) for the asymptotic variance of $\hat{\delta}$.

5. Estimating Effect Size from a Series of Experiments with Small Samples

The method of maximum likelihood produces consistent estimates of the effect size δ as the sample sizes n_i^E, n_i^C become large. If the sample sizes of the k experiments remain small, but the number of experiments, k, becomes large, the maximum likelihood estimator of δ is not necessarily consistent. In this section we consider two alternative estimation procedures which produce consistent estimators as k tends to infinity. One procedure involves linear combinations of unbiased estimators considered in Section 3. The other procedure involves a modification of the likelihood equations to produce a single estimation equation that depends only on δ and sample data.

5.1 Estimation Based on a Linear Combination of Unbiased Estimators

One class of estimators of effect size includes linear combinations of the unbiased estimators of effect size from each experiment. The estimator G from Section 4.1 was an estimator of this type. In Section 4.1 we assumed that each experiment had a large sample size and used a large sample procedure to estimate the optimal weights. In the case of small samples in each experiment, we bypass the step of estimating weights and use the weights derived from an a priori estimate of δ. The easiest estimate of δ to use for calculating weights is $\delta = 0$. Since the optimal weights are given by

$$w_i = \frac{1/v_i}{\sum_{j=1}^{k} (1/v_j)} , \quad i = 1,\ldots k,$$

where v_i is the variance of g_i^U, the variance of g_i^U given in Fact 3 yields weights for $\delta = 0$ as

$$w_i = \frac{(m_i - 2)\tilde{n}_i}{m_i\,[c(m_i)]^2 \sum_{j=1}^{k} \{\frac{(m_j - 2)\tilde{n}_j}{m_j\ c(m_j)^2}\}}, \tag{19}$$

where $m_i = n_i^E + n_i^C - 2$, $i = 1,\ldots k$, and $c(m)$ is given by (4). The weighted estimator

$$G = \sum_{i=1}^{k} w_i g_i^U, \tag{20}$$

where the w_i are given by (19) is an unbiased estimator of δ regardless of the sample sizes of the k experiments for $m_i > 2, i = 1,\ldots k$.

5.2 A Modified Maximum Likelihood Estimator

The method of maximum likelihood produces consistent estimators of δ as the sample sizes n_i^E and n_i^C of each experiment become large. We derive the asymptotic theory for the maximum likelihood estimator in Section 4.2 under the specific assumption that $\pi_i^E = n_i^E/N$ and $\pi_i^C = n_i^C/N$ remain fixed as $N = \sum_{i=1}^{k} (n_i^E + n_i^C)$ tended to infinity. If the sample sizes of each experiment are fixed, but the number k of experiments becomes large, then the method of maximum likelihood does not necessarily produce consistent estimators (Neyman and Scott, 1948). Note that in this later case, the assumptions of Section 4.2 are not met since N tends to infinity but π_i^E and π_i^C tend to zero as k tends to infinity.

In this section we consider estimation based on a modification of the likelihood equations that yields a single estimation equation which depends only on δ and sample data. The logic of the procedure is essentially the same as that proposed by Neyman and Scott (1948) for estimation of a parameter in the presence of indefinitely many incidental parameters.

The estimation equation is a modified form of the likelihood equations. Define the random variable Z_i, $i = 1,\ldots,k$, as the ratio of the between-group sum of squares, $SSBG_i$, to the total sum of squares about the mean, SSY_i, for the i^{th} experiment, $i = 1,\ldots,k$, i.e.,

$$Z_i = \frac{SSBG_i}{SSY_i} \quad , \quad i = 1,\ldots,k \quad . \tag{21}$$

Define the function $\phi_i(Z_i,\delta)$, $i = 1,\ldots,k$, which depends only on Z_i and δ via

$$\phi_i(Z_i,\delta) = \delta Z_i + \sqrt{\delta^2 Z_i^2 + 4(n_i/\tilde{n}_i)Z_i} - 2\delta \quad , \quad i = 1,\ldots,k \tag{22}$$

where $n_i = n_i^E + n_i^C$ and $\tilde{n}_i = n_i^E n_i^C / (n_i^E + n_i^C)$, $i = 1,\ldots,k$. Next define the function $\Phi_i(Z_i, \delta)$ by

$$\Phi_i(Z_i,\delta) = \phi_i(Z_i, \delta) - E\,[\phi_i(Z_i,\delta)] \quad , \quad i = 1,\ldots,k \quad . \tag{23}$$

Note that the expectation $E\,[\phi_i(Z_i,\delta)]$ is a function of δ alone. Given a collection of k experiments and the structural model (1) define the function $F(Z_1,\ldots,Z_k,\delta)$ by

$$F(Z_1,\ldots,Z_k,\delta) = \sum_{i=1}^{k} w_i \Phi_i(Z_i,\delta) \quad , \tag{24}$$

where $w_i > 0$, $\sum_{i=1}^{k} w_i = 1$, $i = 1,\ldots,k$, are weights that are chosen on the basis of sample sizes for the i^{th} experiment. These definitions are used in the statement of the following theorem on a procedure for estimation of δ given by Hedges (1980).

Theorem 4: Let a series of k experiments satisfy the structural model (1) with $0 < |\delta| < d$, where d doesn't depend on k, and where $\{n_i^E + n_i^C \mid i = 1,\ldots,k\}$ is bounded as $k \to \infty$, then the solution, $\hat{\delta}$, to the equation

(25) $$F(Z_1,\ldots,Z_k,\delta) = 0$$

is a consistent estimator of δ as $k \to \infty$, and the estimates have an asymptotic normal distribution (as $k \to \infty$) given by

(26) $$\sqrt{k}(\hat{\delta} - \delta) \sim N(0,\sigma_k^2)$$

where

(27) $$\sigma_k^2 = \frac{\sum_{i=1}^{k} w_i^2 E\ \phi_i^2(Z_i,\delta)}{\sum_{i=1}^{k} w_i E\left[\frac{\partial \phi_i(Z_i,\delta)}{\partial \delta}\right]}^2 .$$

Proof: A complete proof is given in Hedges (1980).

It is obvious that the estimate cannot be obtained in closed form. Unfortunately the expectations in the estimation equation and the asymptotic variance σ_k^2 cannot be obtained in closed form either. The expectations however can be obtained in terms of hypergeometric functions. For the estimation equation we need

$$E\ [\phi_i(Z_i,\delta)\ ] = \delta\ E(Z_i) + \delta E\ [\sqrt{Z_i}\ \sqrt{Z_i + c_i}] - 2\ \delta,$$

where Z_i has the noncentral beta distribution with parameters ½ and $(n_i^E + n_i^C - 2)/2$ and noncentrality parameter $\tilde{n}_i\delta^2/2$, and $c_i = 4(n_i^E + n_i^C)/\tilde{n}_i\delta^2$.

The asymptotic variance depends on $E\ [\ \phi_i^2(Z_i,\delta)]$ and $E\ [\ \partial\phi_i(Z_i,\delta)/\partial\delta]$ which are more complicated. Hedges (1980) showed that

$$E[\Phi_i^2(Z_i,\delta)] = \delta^2\{2E(Z_i^2) + c_iE(Z_i) - [E(Z_i)]^2$$

$$+2E[Z_i^{3/2}\sqrt{Z_i+c_i}] - 2E[\sqrt{Z_i}\sqrt{Z_i+c_i}]E(Z_i) - [E(\sqrt{Z_i}\sqrt{Z_i+c_i})]^2\},$$

where c_i and Z_i are defined as above. Hedges also gave the following expression for $E[\partial\Phi_i(Z_i,\delta)/\partial\delta]$

$$E\left[\frac{\partial\Phi_i(Z_i,\delta)}{\partial\delta}\right] = \delta^2\{E(Z_i) - E(W_i) + E(\sqrt{Z_i}\sqrt{Z_i+c_i})$$

$$-E(\sqrt{W_i}\sqrt{W_i+c_i})\},$$

where Z_i and c_i are defined as above and W_i is a noncentral beta variate with parameters 3/2 and $(n_i^E + n_i^C - 2)/2$ and noncentrality parameter $\tilde{n}_i\delta^2/2$.

The series expression for the noncentral beta density $f(x;a,b;\lambda)$ with parameters a and b and noncentrality parameter λ is well known (see e.g. Rao, 1965) as

$$f(x;a,b;\lambda) = e^{-\lambda}\sum_{r=0}^{\infty}\frac{\lambda^r}{r!}\, f(x;a+r,b;0),$$

where $f(x;a,b;0)$ is the central beta density. Using this expression for the density we obtain

$$E(Z) = e^{-\lambda}\sum_{r=0}^{\infty}\frac{(a+r)}{(a+b+r)}\,\frac{\lambda^r}{r!},$$

$$E(Z^2) = e^{-\lambda}\sum_{r=0}^{\infty}\frac{(a+r)}{(a+b+r+1)(a+b+r)^2}\,\frac{\lambda^r}{r!}, \qquad (28)$$

where Z is a noncentral beta variate with parameters a and b and noncentrality parameter λ. Evaluation of the expectation

(30) $$E\left[Z^{\alpha} \sqrt{Z + c} \right]$$

is more involved. We evaluate (30) for each term of the series, i.e. for central beta variates. To evaluate (30) for central beta variates we need the hypergeometric series representation from Erdelyi (1954, p.186)

$$\int_0^1 x^{d-1}(x + c)^{e} \quad (1-x)^{h-1} \, dx = c^{e} \, B(h,d) \, {}_2F_1(-e,d;d + h;-1/c$$

where $B(h,d) = \Gamma(h)\Gamma(d)/\Gamma(h + d)$, and ${}_2F_1(\alpha,\beta;\gamma,x)$ in Gauss's hypergeometric function,

$${}_2F_1(\alpha,\beta;\gamma;x) = \sum_{r=0}^{\infty} \frac{\Gamma(\alpha + r) \, \Gamma(\beta + r)\Gamma(\gamma)}{\Gamma(\alpha) \, \Gamma(\beta)\Gamma(\gamma + r)} \frac{x^r}{r!} .$$

Therefore, since the r^{th} term of the series for the expectation (30) is $e^{-\lambda}\lambda^r/r!$ times

$$\int_0^1 x^{\alpha} \sqrt{x + c} \; \frac{\Gamma(a_r + b)}{\Gamma(a_r) \, \Gamma(b)} \; x^{a_r - 1} \, (1-x)^{b-1} \, dx,$$

where $a_r = a + r$, the r^{th} term can be represented as a hypergeometric series. Thus the final representation for $E\left[Z^{\alpha} \sqrt{Z + c} \right]$ is

(31) $$c^{1/2} e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} \frac{\Gamma(a_r+b)\Gamma(a+\alpha)}{\Gamma(a_r) \, \Gamma(a_r+b+\alpha)} \; {}_2F_1 \, (-1/2, a_r + \alpha; a_r + b + \alpha; -1/c),$$

where $a_r = a + r$. Equations (28), (29), and (30) provide the representation of all expectations needed for the estimation equation (25) and the expression (27) for the asymptotic variance.

5.3 Comparing Estimators of Effect Size from a Series of Small Experiments

The modified maximum likelihood estimator given in Section 5.2 is not a serious competitor to the estimators based on linear combinations of unbiased estimators. Although the modified maximum likelihood estimator (25) is consistent (as $k \to \infty$), it is computationally demanding and it yields estimates with larger asymptotic variance than the weighted estimator (20). Since both estimators (20) and (25) are consistent (as $k \to \infty$), one way to compare the two estimators is to compare their variances. When all the experiments in a series have equal sample size, i.e. when $n_i^E = n_i^C = n$, $i = 1,\ldots k$, the ratio of the variances of (20)and (25) are independent of k. Thus we can easily compute the relative efficiency of the two by computing the ratio of the variances of the two estimators. Table 2 is a presentation of the relative efficiency of the two estimators for various sample sizes and effect sizes. The superiority of the linear combination of unbiased estimates is very clear.

Table 2

Relative Efficiency σ_1^2/σ_2^2 of the Modified Maximum Likelihood Estimator with Asymptotic Variance σ_2^2 to the Unbiased Estimator with Variance σ_1^2 for Sample Size n per Group and Effect Size δ

n	$\delta = .20$ σ_1^2/σ_2^2	$\delta = .50$ σ_1^2/σ_2^2	$\delta = 1.0$ σ_1^2/σ_2^2
3	.868	.745	.614
4	.649	.495	.306
5	.526	.364	.198
6	.440	.281	.140
7	.377	.223	.104
8	.328	.182	.081
9	.289	.151	.065
10	.257	.131	.053

6. Testing Whether a Series of Experiments Share a Common Effect Size

The methods of estimating effect size from a series of experiments were all derived under model (1) which implies that all experiments share a common effect size. In practice it is important to see if this assumption of homogenous effect sizes is realistic. In this section we present two natural statistics for testing the hypothesis that the population effect sizes from a series of k experiments are equal, and show that these two statistics have the same asymptotic distribution.

Formally each of these tests involves a hypothesis about the model

$$Y_{ij}^E = \beta_i\gamma_i + \beta_i\delta_i + \varepsilon_{ij}^E, \quad i = 1,\ldots,k, \; j = 1,\ldots n_i^E,$$

(32)

$$Y_{ij}^C = \beta_i\gamma_i + \varepsilon_{ij}^C, \quad i = 1,\ldots,k, \; j = 1,\ldots,n_i^C,$$

where ε_{ij}^E and ε_{ij}^C are distributed independently as $N(0,\beta_i^2)$, $i = 1,\ldots,k$. A test of homogeneity is a test of the hypothesis

$$H_o: \delta_i = \delta, \; i = 1,\ldots,k,$$

versus the alternative that at least one of the δ_i is not equal to the others. Note that model (32) under H_o is exactly model (1).

6.1 The Likelihood Ratio Test for Homogeneity of Effect Size

One approach to testing the homogeneity of the δ_i's is the likelihood ratio test comparing a model with all δ_i's equal versus a model which fits a different δ_i to each experiment in the series. Define the likelihood L_i for the i^{th} experiment as in (11). Simplification of (11) yields L_i as a function of $\bar{Y}_i^E, \bar{Y}_i^C$, the total sum of squares about the mean of the i^{th} experiment SSY_i, and the parameters β_i and δ. The resulting expression is

$$L_i(\bar{Y}_i^E, \bar{Y}_i^C, SSY_i, \beta_i, \delta) = \tfrac{1}{2} \ln(2\pi)^{n_i} - n_i \ln(\beta_i)$$

$$- \frac{1}{2\beta_i^2} \{SSY_i + \tilde{n}_i \beta_i^2 \delta_i^2 - 2\tilde{n}_i \beta_i \delta (\bar{Y}_i^E - \bar{Y}_i^C)\}, \quad i = 1, \ldots, k,$$

where $n_i = n_i^E + n_i^C$, $\tilde{n}_i = n_i^E n_i^C / (n_i^E + n_i^C)$, $i = 1, \ldots, k$, and SSY_i is the total sum of squares about the mean for the i^{th} experiment.

The test of statistic for the likelihood ratio test of homogeneity of effect size is

$$X = 2\sum_{i=1}^{k} \{L_i(\bar{Y}_i^E, \bar{Y}_i^C, SSY_i, \sqrt{n_i/(n_i - 2)}\, S_i, \hat{\delta}_i) - L_i(\bar{Y}_i^E, \bar{Y}_i^C, SSY_i, \hat{\beta}_i, \hat{\delta})\}, \tag{33}$$

where $\hat{\delta}_i$ is the maximum likelihood estimate of the i^{th} experimental effect size given by (6), S_i^2 is the usual unbiased estimate of the within group variance of the i^{th} experiment, $\hat{\beta}_i$ is the maximum likelihood estimate of β_i under H_o, and $\hat{\delta}$ is the maximum likelihood estimate of δ under H_o.

Theorem 5: Suppose a series of k experiments under model (32) all have large sample sizes, then the statistic X defined by (33) is distributed approximately as a chi-square under H_o with (k-1) degrees of freedom.

6.2 An Alternative Test for Homogeneity of Effect Sizes

One disadvantage of the likelihood ratio test of homogeneity of effect sizes is that it requires calculation of the maximum likelihood estimate of δ under H_o, which must be obtained numerically. In this section we obtain a computationally simpler test for homogeneity of effect size. If the sample sizes n_i^E, n_i^C are large in each study, then the distribution of each g_i is approximately normal, i.e.

$$g_i \sim N(\delta, \sigma_i^2(\delta))$$

where $\sigma_i^2(\delta)$ is given by (5). Therefore the statistic

$$(g_i - \delta)^2 / \sigma_i^2(\delta)$$

is distributed as a chi-square with one degree of freedom. This principle is extended to generate a test for homogeneity given in the following theorem.

Theorem 6: Suppose k experiments under model (32) all have large sample sizes n_i^E, n_i^C, then the statistic H, defined by

$$H = \sum_{i=1}^{k} \frac{(g_i - G)^2}{\sigma_i^2(g_i)} ,$$

has an approximate chi-square distribution with (k-1) degrees of freedom under H_o, where G is the estimator (9) and $\sigma_i^2(\delta)$ is given by (5).

Proof: This is an application of a well known theorem on weighting estimators (see e.g. Rao, 1965, p.389-390).

Note that the two statistics X and H both have the same asymptotic dis-

tribution, but H is considerably easier to calculate. The normal approximation used (implicitly) in the derivation of the distribution of H has also been extensively studied (see e.g. Johnson & Welch, 1939). The accuracy of this approximation in moderate samples suggests that the large sample expression for the distribution of H may be quite good even in samples of moderate size.

7. Example

Some of the techniques described in this paper were applied to studies in research on teaching by Hedges and Gage (1981). Some data from five studies that examined the effects of open versus traditional teaching on measures of student cooperativeness are given in Table 3. For each study, the sample sizes and estimates using each of the four estimators are given. Note that the sample sizes of each study are relatively large and that the four estimators yield effect size estimates in each study that do not differ to any large degree. The homogeneity statistic H has the value H = 8.772. Comparing this value with the 95% critical value of the chi-square distribution with 4 degrees of freedom (9.49), we see that we cannot reject the hypothesis of homogeneity of effect sizes for these five studies. For comparison we calculated the likelihood ratio test statistic X for homogeneity of effect sizes and obtained the value X = 14.00.

Since the studies all have relatively large samples we compute both the maximum likelihood estimate of effect size and the weighted estimator of effect size (9) with weights estimated from the samples. The estimated variance $\hat{\sigma}^2(g)$ of the effect size and the estimated weights $\hat{w}$ for each study are given in Table 3. The weighted overall estimate of effect size is G = .118 and an asymptotic 95% confidence interval for δ based on G and Theorem 2 is

$$-.129 \leq \delta \leq .365.$$

The maximum likelihood estimate $\hat{\delta}$ of δ based on data from the five studies is obtained by solving (13) numerically. The value of $\hat{\delta}$ obtained is $\hat{\delta} = .107$ and an asymptotic 95% confidence interval for δ based on $\hat{\delta}$ and Theorem 3 is

$$-.141 \leq \delta \leq .355 .$$

Table 3

Estimates of Effect Size from Five Studies along with Sample Sizes, Estimated Variance of the Estimators $\hat{\sigma}^2(g)$, Estimated Weights $\hat{w}$ Assigned to Each Study

	Open N	Trad N	$\hat{\delta}$	g	g^U	g^S	$\hat{\sigma}^2(g)$	$\hat{w}$
1	16	11	.492	.473	.459	.449	.157	.081
2	30	30	.187	.183	.181	.179	.067	.191
3	30	30	-.537	-.528	-.521	-.516	.069	.185
4	44	40	.099	.098	.097	.096	.048	.267
5	37	55	.433	.429	.425	.423	.046	.276

8. Conclusions

The results of Section 3 suggest that there are definite differences among the four estimators of effect size for a single experiment with a small sample size. The maximum likelihood estimator and Glass's estimator have little to recommend them. The unbiased estimator is easy to calculate and it is more precise than either Glass's estimator or the maximum likelihood estimator. If considerations of bias outweigh considerations of accuracy (mean squared error) then the unbiased estimator might be chosen over the shrunken estimator. If the accuracy of the estimator is the most crucial factor, the shrunken estimator has a clear advantage over the other three estimators in small samples. If the sample size of the experiment is large, all four estimators will produce very similar estimates.

When estimates of effect size are desired from a series of experiments with large samples, the results of Section 4 suggest that either weighted estimates or the method of maximum likelihood will yield estimators with the same asymptotic distribution. Calculation of the maximum likelihood estimate of effect size from a series of experiments requires iterative solution of the likelihood equation. Hence the weighted estimator is much easier to calculate and may be preferrable for the sake of simplicity.

In the case of estimation of effect size from a series of experiments with small sample sizes, linear combinations of unbiased or shrunken estimators are probably preferable. The alternative estimator (25) is not only less precise, but involves considerable computation.

In any situation involving estimation of a common effect size from a series of experiments, the homogeneity of effect size across studies should be investigated. Two statistical tests for homogeneity of effect sizes were

given. Both test statistics have the same asymptotic distribution, but the statistic H is clearly much easier to calculate. It is the experience of the author that many series of studies which investigate the same question fail tests of homogeneity of effect size by a wide margin. If a series of studies cannot credibly be said to share a population effect size, then point estimates of "the" (single) population effect size are difficult to interpret. Such analyses may lead to oversimplifications of a complex situation in which different subgroups of studies have different population effect sizes.

9. References

Erdelyi, A. Tables of Integral Transforms, Volume II. New York: McGraw-Hill, 1954.

Glass, G.V. Primary, secondary, and meta-analysis of research. Educational Researcher, 1976, 5, 3-8.

Glass, G.V & Smith, M.L. Meta-analysis of the relationship between class-size and achievement. Educational Evaluation and Policy Analysis, 1979, 1, 2-16.

Hedges, L.V. Combining the Results of Experiments Using Different Scales of Measurement. Unpublished doctoral dissertation, Stanford University, 1980.

Hedges, L.V. Distribution theory for Glass' estimator of effect size and related estimators. Journal of Educational Statistics, 1981 (in press).

Hedges, L.V. & Gage, N.L. The empirical evidence on the effects of open education. Final Report of the Stanford Research Synthesis Project, volume II. Stanford University School of Education; Stanford, California, 1981.

Johnson, N.L. & Welch, B.L. Applications of the noncentral t-distribution. Biometrika, 1939, 31, 362-389.

Neyman, J. & Scott, E.L. Consistent estimators based on partially consistent observations. Econometrica, 1948, 16, 1-32.

Rao, C.R. Linear Statistical Inference and Its Applications. New York: John Wiley, 1965.

Smith, M.L. & Glass, G.V. Meta-analysis of psychotherapy outcome studies. American Psychologist, 1977, 32, 752-760.

Stein, C. Inadmissability of the usual estimator for the mean of a multivariate normal distribution. In Proceedings of the third Berkely Symposium on mathematical statistics and probability,1. Berkely, California: University of California Press, 1956.